DOCUMENT RESUME

ED 273 779 CE 044 932

AUTHOR Campbell, Clifton P.
TITLE Criterion Referenced Performance Tests for Measuring Task Proficiency.
PUB DATE Sep 86
NOTE 47p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Concurrent Validity; *Criterion Referenced Tests; Evaluation Criteria; *Performance Tests; Predictive Validity; Scores; Secondary Education; Student Evaluation; *Test Construction; *Testing Problems; *Test Use; *Vocational Education

ABSTRACT

This guide explains the benefits of and procedures for constructing and administering criterion-referenced performance tests in vocational education programs. Addressed in the individual sections are the following topics: the rationale for using task performance measures in vocational education; the predictive validity and physical fidelity of criterion-referenced performance tests; simulation; unitary and multiple tasks; development of a task performance measure; constraints of criterion-referenced performance testing (cost, equipment and facilities, time, personnel, and other constraints); measurement of products and/or processes; procedures for scoring (product and process ratings, checklists, rating scales, and establishment of cut-off scores); and validation. An appendix contains a sample task performance measure for use in a course to train chemistry technicians, a task performance measure worksheet, and sample evaluation criteria for a task performance measure. (MN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED273779

Criterion Referenced Performance Tests for Measuring Task Proficiency

Clifton P. Campbell

Technological and Adult Education Department

The University of Tennessee

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C P Campbell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

September 1986

CE 044932

TABLE OF CONTENTS

Page

LIST OF TABLES....iv

LIST OF FIGURES.... v

Introduction.... 1

Rationale.... 1

Predictive Validity.... 3

Physical Fidelity.... 4

Simulation.... 5

Unitary and Multiple Tasks.... 5

Developing a TPM.... 7

Testing Constraints.... 8

- Cost.... 9
- Equipment and Facilities.... 9
- Time....10
- Personnel....10
- Other Constraints....11

Measure Product and/or Process....11

Determine Conditions....12

Determine Cues....14

Determine Standards....15

Testing All or Part of the Task....17

Documentation....18

Page

Scoring Procedures.....19

Product Rating.....20

Process Rating.....20

Checklists.....21

Rating Scales.....21

Establishing Cut-off Scores.....24

Validation.....24

Tryout and Revise.....25

A Comment.....28

REFERENCES.....29

APPENDIX.....32

LIST OF TABLES

Table

1. Task Performance Measure Documentation

2. Performance Measurement by Means of Product or Process

LIST OF FIGURES

Figure

1. Task Performance Measure Worksheet

2. Developing Task Performance Measures

3. Checklist for Rating Instrument Flying Proficiency

4. Descriptive Scale for Rating Pilot Landing

5. Forms of a Graphic Scale

Introduction

Publicly supported vocational education programs are frequently not held in the highest regard by the community they seek to serve. Clearly the image needs to change! Additionally, the profession itself is concerned about the discouraging outcomes of some vocational programs. Governing bodies, employers and taxpayers are all insisting that vocational programs become more accountable.

Accountability, the driving force behind performance based education/training, is substantially increased when Task Performance Measures are used. This is the case because they provide the means for basing instruction and evaluation on job performance requirements rather than arbitrarily derived subject matter.

A task performance measure evaluates task proficiency using actual equipment under work conditions and standards. Ideally it is conducted in the job environment. Task performance measures are a means of keeping instruction consistent with job requirements and they clearly distinguish whether or how well an individual can perform.

Rationale

Once it has been determined which job tasks will be included in an instructional program, it is possible to develop performance measures. The task performance measure (TPM), as the name indicates, involves actual

performance (doing) by an examinee. It is a criterion referenced performance test, prepared at the task level, which provides a direct and realistic method for measuring task proficiency.

With this type of test, complex behaviors, such as those involved in operating or repairing equipment can be evaluated in an objective and meaningful way.

On the basis of TPM feedback, an instructional program can be revised to improve its effectiveness. Thus, an important aspect of TPMs is that they can be used in assessing program effectiveness as well as trainee competence. TPM's are beneficial to vocational programs because they:

1. distinguish between those who can meet attainment standards when performing a job task from those who cannot, thereby controlling program output.
2. serve as a keystone for the development and control of programs with explicit learning objectives and appropriate content.

In addition, TPM's are of benefit to employers because they:

1. eliminate unnecessarily restrictive standards in position (job) descriptions.
2. avoid the basis of real or apparent discrimination in training, employment, and promotion practices.
3. help assure, when used as employment and promotion exams, that the work force consists of the most appropriate individuals for doing the job.

Maximizing the uses of TPMs assures a degree of compatibility in training, evaluation, and related personnel practices. This multiple use, in addition to the previously identified advantages, justifies the investment required to construct, validate, and administer TPMs.

Predictive Validity

TPMs are criterion referenced tests and as such should possess high predictive validity; that is, performance on the test should be highly related to performance of the task. A TPM is said to have high predictive validity if those who pass are those who perform the task well. The ideal TPM would have perfect predictive validity! Without exception, those who passed could perform the job task, and all those who failed could not.

EXAMPLES

1. One of the tasks of a Motor Transport Operator is to fill out forms associated with vehicle operation and maintenance. The TPM for this task is identical to the task itself. In this case, predictive validity would be very high.
2. Another task of a Motor Transport Operator is to drive a truck with a trailer from a specified location to a specified destination. Task performance requires driving under a variety of conditions. The TPM for this task is:

   Behavioral Action: Drive truck with trailer,

Conditions:	Standards:
a. over a ditch	without getting stuck
b. over a muddy road	without getting stuck
c. down a steep and winding grade	using proper gears and brake action
d. up a steep and winding grade	using proper gears

In the second example, the predictive validity probably would still be reasonably high. Examinees who pass the TPM probably could perform the task. Those who fail probably could not perform the task.

When it is practical to measure the task, in the same way that it is performed on the job, it is possible to determine the predictive validity of the TPM. In such cases, the primary objective in developing TPMs is to achieve the highest possible predictive validity, while at the same time, keeping testing costs, time, and safety within acceptable limits.

There are two situations where paper and pencil tests are particularly suitable for measuring performance. The first concerns those tasks that are accomplished by using paper and pencil in the "real world," e.g., filling out forms associated with vehicle operation and maintenance. Because paper and pencil are used in "real world" task performance, the TPM should likewise be a paper and pencil test.

The second situation where a paper and pencil test would be appropriate is when mental discrimination is the most important aspect of the task. For example, select from a number of routes the one that is best for driving from one location to another, based upon the conditions provided.

Physical Fidelity

Physical fidelity refers to how well the behavioral action, conditions, cues, and standards of the TPM approximate those of the task. In example one, filling out forms associated with vehicle operations and maintenance, the TPM has the highest possible physical fidelity. This is because TPM performance is identical to task performance. The second example, drive a truck with a trailer, has lower but still reasonably high physical fidelity.

With both of these examples, predictive validity is more important than physical fidelity. If a paper and pencil test, with lower physical fidelity, could be shown to have practically the same predictive validity, and if this approach were preferable in terms of cost, time and safety, it should be used.

Unfortunately, having high physical fidelity does not ensure that predictive validity will be high. However, if predictive validity cannot be used, one must often settle for high physical fidelity.

Simulation

Broadly defined, simulation is any change from reality, an imitation. If a TPM cannot be administered under "real world" conditions, some form of simulation may be used. When deciding whether or not to use simulation, the following factors should be considered: (a) the effect on productivity when actual equipment and/or facilities are used for testing, (b) potential dangers to personnel or damage to equipment, and (c) the cost of using materials, supplies, and equipment under work conditions.

Unitary and Multiple Tasks

Another factor that affects construction of TPMs is whether the task is unitary or multiple. A unitary task is one that is always performed in exactly the same way with exactly the same inputs. A multiple task is one that has a number of possible inputs.

There are two types of multiple tasks. One types is always performed by following basically the same procedure.

EXAMPLE

A task requiring multiplication of three-digit numbers by three-digit numbers is a multiple task. The multiplication is always performed by following basically the same procedures. However, there are 810,000 possible inputs (combinations of the two three-digit numbers.)

In this example, it would be impossible to test all possible inputs to make sure an individual knows how to do them all. For unitary tasks, the TPM can measure the total: however, for multiple tasks a representative sample must be used.

The second type of multiple task is one in which the inputs vary, and the task is performed differently depending on the input. This means the input is a cue that determines the appropriate response.

EXAMPLE

A task performed by a police officer is to apprehend suspects. The procedure for performing this task depends heavily on the inputs (cues). Answers to questions like the following will determine how the task is performed.

- Is the suspect armed or not?
- Is the suspect on drugs or not?
- Is the suspect in a vehicle or on foot?
- Is the suspect alone,with a few others, or in a crowd?
- Is the police officer alone or with others?
- Is the police officer in a vehicle or on foot?

As with the first type of multiple task, the TPM for this task cannot measure the total task but must test a sample of the possible variations. However, for this type of task, the TPM must do more than measure the adequacy of the performance. It also must measure the appropriateness; that is, whether the performance was the correct response to the particular inputs (cues).

Developing a TPM

The makeup or composition of a TPM may vary to suit a particular need. A useful format is provided by the TPM Worksheet shown in Figure 1. This worksheet has blank spaces to fill in: (a) a statement of behavioral action to be taken; (b) the performance conditions under which the test is to be administered, including cues that indicate the need for particular responses; and (c) the attainment standards which are a measure of the adequacy of performance; and other key information. Additionally, space is provided for the task elements and/or steps that are the actions which make up the total performance.

Place Figure 1 about here

It is not always feasible to duplicate every aspect of a task in a TPM. For this reason, the procedure for developing a TPM is basically one of (a) considering in what ways the test must be different from the task, because of various testing constraints, and (b) constructing a test that is the best compromise.

Figure 2 shows, in flow chart form, the procedural steps in developing and validating a TPM.

The recommended approach to following this nine-step procedure is to take one task at a time and determine (a) testing constraints, (b) if product, process, or both should be measured, (c) conditions, (d) cues, (e) standards, and (f) if all or part of the task will be tested. Then develop a sampling plan and construct the TPM. Take the next task and repeat the procedure.

The first determinations should be considered tentative. After working through a few tasks it is advisable to go back and make revisions and changes as appropriate. When all of the TPMs and scoring procedures have been developed, they should be validated or verified.

Place Figure 2 about here

Testing Constraints

If there were no constraints of (a) money, (b) equipment and facilities, (c) time, (d) personnel, and (e) other resources, every TPM would mirror the task it is intended to measure. That is, the TPM would consist of observing the job incumbent or examinee performing the task in the job environment and noting whether or not the standards were attained. For example, the TPM for "rotate tires on a truck" ought to be identical to task performance. However, since constraints often exist, the TPM may be different from the task.

The first step in developing TPMs that have high predictive validity, high physical fidelity, or both, is to review the task to determine testing

constraints. The following constraints can force a change from a higher to a lower physical fidelity TPM. It should be noted that these constraints are for the most part interrelated. For example, cost, the availability of equipment and facilities, time, and personnel availability, are often all different aspects of the same problem.

Cost. The first constraint is one of economics. Costs are an important factor in developing and administering any TPM and they must be kept within limits.

EXAMPLE

> In most cases it would be cost prohibitive to have a demolition specialist destroy a building in order to measure task proficiency. There are other more practical means of testing this task. If the task specifies demolishing a building, the TPM may need to be modified so that the building is not actually demolished, but the job incumbent or examinee demonstrates the process and procedures up to but not including the actual demolition. Simulation is another alternative that should be considered.

Equipment and facilities. Often equipment and/or facilities are not continuously available for test administration. This is especially true for sophisticated equipment and specialized facilities. An example of a constraint concerning equipment availability might involve a course on computer trouble-shooting; computer downtime may be so costly as to negate its use for testing purposes.

When testing would tax equipment and/or facilities beyond feasible limits, it may be necessary to select for testing only those tasks which would cause the least inconvenience. The remainder of the tasks might be tested using simulation.

Time. The practical constraint of time availability is easily understood. Often the situation is such that it is impractical to test the entire task in the time available.

EXAMPLE

Perhaps the job task requires hiking six miles over hills during inclement weather. Since this task would take too long to test in most situations, it must be modified to permit testing in less time.

In general, some time constraint must be placed on test administration. This and the time standard for task performance limit the amount of time that can be spent on each TPM. If performance of some of the tasks requires more time than is available for testing, the TPM should be constructed using a sample of the critical task elements.

Personnel. The availability of sufficient people may also impose a constraint on task performance. This can occur when more than one individual is required to perform a task under normal conditions in the job environment.

EXAMPLE

If, for example, it takes two individuals to operate a machine, and only one examinee is available for the TPM, someone trained in the other machine functions will be required in order to administer the test. If the second individual, an incumbent worker or another trainee, is not available, there will be insufficient personnel to conduct the test under normal operating conditions.

In this example, it would be preferable to schedule the TPM for two examinees simultaneously. Although individual performance is usually tested, multiple or even team performance can be measured. In such cases, however, more than one observer may be necessary to ensure effective scoring.

Other constraints. There are other less common constraints which may be encountered in the development of a TPM. These constraints include:

1. logistics
2. supervisory effectiveness
3. communications
4. ethical considerations
5. legal considerations

It should be remembered that in most cases constraints are interrelated. The practical constraint in the example of computer troubleshooting was categorized under equipment availability, but it could also be categorized under costs.

Considering the limitations of cost, equipment and facilities, time, personnel, space, safety requirements and other factors, it is obvious that the highest possible physical fidelity is not always practical or even desirable. In such cases, the TPM developed must be viewed as the best possible trade-off with reality.

Measure Product and/or Process

The task standard can be based on a product, a process, or both. The end product is the most obvious task output; it can be observed and inspected. A computer program, or the results of a desk audit of an initial draft computer program are products of task performance.

The other output which signals task performance is the completion of the process involved. The process itself may leave no record. The skills involved (for instance, in safe driving) can best be evaluated by close observation.

In some cases, both product and process must be examined as task output. Many tests evaluate processes which result in products in order to provide feedback on process errors which affect the product or related policies and practices.

In certain cases, the processes in performing a task may be critical in that they insure personnel safety or prevent equipment damage. For instance, a driver may arrive at destination B from point A as required by the task, but, in the process, violate traffic laws and safety regulations. In such cases, product evaluation by itself would not be adequate.

The TPM will usually measure the same factors, product, process, or both, used as a basis for the task standard; however, this is not an essential practice. If considerable simulation will be required, it may be more practical to measure another factor.

EXAMPLE

> The standard for the Demolition Specialist's task of "demolish a building" might be that the building is completely leveled without personal injury. The TPM would probably measure the <u>procedure</u> used rather than the <u>product</u>.

<u>Determine Conditions</u>

Conditions are prerequisite requirements that exist for task performance. They include the: (a) setting; (b) tools, equipment and

supplies; (c) personnel; (d) references; and (e) other considerations such as safety, needed to accomplish a task. Conditions may be categorized as constant or variable.

EXAMPLE

A land navigation task would be to hike from point A to point B on foot. The constant conditions are:

A map of the area

A compass

The variable conditions are:

All types of weather - dry, rain showers, thunderstorm, snow, clear, high wind, gusty winds, no wind, chilly, frost, hail, below freezing, cloudy, sunny, warm

Across all types of terrains - plains, deserts, forests, valleys, hills, mountains.

At any time - daytime, nighttime, with stars, no stars, moon, no moon.

If the variable conditions (specifying the weather, terrain and time of day) had been left out, the TPM would be somewhat imprecise. Yet it is impossible to combine all the variable conditions especially when factors such as the weather cannot be controlled.

When testing constraints preclude using all the variable conditions identified, the TPM developer is forced to use a sample of them. Those conditions that have significant effects on the ability to perform under all the conditions should be included.

Determine Cues

Often TPM cues must be different from those that prompt or signal a job incumbent to perform a specific task. This is frequently the case even though it reduces the fidelity.

EXAMPLE

The cue for a nurse to perform mouth-to-mouth resuscitation would be that the victim was not breathing. However, this is an unrealistic cue for the TPM. While changing the cue reduces the fidelity of the TPM, probably the best that could be done would be to use a cue such as having the examiner say, "the victim lying face down on the floor is unconscious, breathing has stopped, take immediate action."

Insofar as constraints permit, the critical cues should be realistic. A critical cue is one for which the proper response determines success or failure in performing the task.

EXAMPLE

The task of "repair malfunctioning equipment" could be such that the actual repair is relatively simple when the examinee knows what to fix. The critical part of the task might be determining what needs fixed. In this example, if the TPM cue were having the examiner say, "Assume the equipment needs a new switch," the TPM would have very low fidelity and probably very low predictive validity.

Determine Standards

The criteria for TPM attainment standards can include: (a) quality (degree of excellence requirements), (b) quantity (rate of production requirements), (c) time limit (speed), and (d) the amount of supervision or assistance to be provided. These standards are applied consistently to all examinees so that each individual is rated like all others.

TPM standards may relate to the process or product of performance or a combination of both. Process standards usually include time limits and the required step-by-step procedure to be followed. Product standards describe the tangible end-product required and criteria for judging that product.

Attainment standards should be transferred from job analysis information to the TPM whenever possible. The introduction of artificial standards, especially time limits, must be avoided.

On occasion the test designer or examiner becomes the source of an artificial performance standard. This generally happens when there is no verifiable "real-world" requirement to perform a task correctly within a specified time.

EXAMPLE

> The task "change a flat tire" may be required of a driver. However, if the condition of "mountainous terrain" and the standard "within five minutes," are added the test may have been used to set the standard when there was no standard implied in the task. It might be troublesome if the driver required an hour to perform the task, but what if it only took ten minutes? To the statistician that means being off by a factor of 100%, but to the driver, it may

> be only a minor problem. Unless there are field requirements for conditions and standards, they should ordinarily be stated to the minimum acceptable requirement. In the case of changing a flat tire, it may be necessary to do a time and motion study just to discover a five-minute procedure, when almost anyone could do it in ten minutes.

Unnecessarily high attainment standards are costly in testing efforts and of dubious value in decision making.

Standards relating to procedures require that certain steps be followed in the performance of the task. If certain steps are omitted or done in the wrong sequence when sequence is important, then the examinee does not meet the standard. The idea behind the use of such a standard is that there is no such thing as performing some skills half-right.

Some performance measures may include a margin of error. For example, a performance measure in land navigation may allow a scout to plot an azimuth within a tolerance of plus or minus two degrees. If the answer falls within that range, the standard has been met. Other tests may allow an examinee to vary the sequence in which certain steps of a procedure are performed. This should be the case when the sequence of step performance is not important.

Some standards include requirements related to the final product. Occasionally, procedures followed in turning out the end-product may be less important than the quality of the product itself. However, if procedures are followed correctly, the chances that the product will meet the standards should be improved.

Another class of performance standards are highly qualitative, such as those for briefings and research studies. The effectiveness of an oral

briefing is difficult to break down into elements that can be objectively scored on a right-wrong basis. Evaluation of its worth is essentially judgmental and qualitative rather than quantitative.

Intellectual and qualitative task performances must be evaluated carefully. This is critical in order to avoid measuring only the less important parts of the task that are more easily scored. Even though evaluation of these tasks is mostly judgmental and qualitative, a quantitatively measurable standard for performance measurement purposes must be derived.

Testing All or Part of the Task

The decision to test all or part of the task depends largely on whether it is a unitary or a multiple task. For unitary tasks, that is, tasks that have only one doctrinally correct procedure, all of the task is tested.

The action, conditions, cues, standards, and constraints previously identified provide guidance in determining what can be done realistically. If the whole task can be tested, it should be. If not, a part-task test that will accurately predict job performance should be developed.

When there are multiple possible actions, conditions, or cues, it must be decided if they are all equally likely to occur and if there is the same criticality involved with the occurrence of each one. If they are equal, each one should be included in an alternate form of the TPM. If the likelihood of occurrence or the criticality is unequal, then a sampling plan should reflect that.

Alternate forms are equal but different versions of the TPM. Constructing alternate forms makes it possible to include all of the important variables without making any single test too long to administer. Occasionally, it is not practical to construct alternate forms because of the large number of variables involved.

EXAMPLE

If the task required multiplying three-digit numbers by three digit numbers, there would be 810,000 possible combinations of the two three-digit numbers. It would be impractical to construct enough alternate forms of the TPM to include all of these problems.

A predetermined quantity of randomly selected numbers could be used for the TPM. In that way, no one would know, until the testing began, which numbers would be selected.

This raises a critical question: Why not initially select three or four problems such as:

987 x 639 = ? 542 x 861 = ?

609 x 713 = ? 484 x 275 = ?

and let these be the TPM? The reason is that one of the important uses of TPMs is to serve as a basis for the development and control of training. This means the training program would be designed to teach people to pass only the four TPM problems. Even if other problems were taught, there would be no way to know they had been learned since they are not in the TPM.

Documentation

Documenting a TPM is quite similar to documenting a task. Like tasks, TPMs have conditions and standards. They also have procedural steps that often are similar to elements. Table 1 is an example of TPM documentation. By attaching TPM documentation to the corresponding task documentation, a complete and useful package of information is made available for training program design and development.

Place Table 1 about here

Scoring Procedures

Scoring procedures used during the measurement of task performance should be reliable and unbiased. Reliability means that evaluation of task performance will not vary over time, and different examiners will evaluate the examinee the same way in any given testing situation. Unbiased scoring procedures ensure that evaluations are not influenced by inappropriate factors. Reliable and unbiased scoring procedures should include clear, measurable standards for evaluating task performance and call for the use of examiners skilled in administering TPMs. The key principle to observe in scoring is objectivity. This is achieved by setting precise standards and training the examiner to apply them.

Some tasks require the examinee to perform a series of procedural steps. An observer is required to record the tested performances on each essential step in order to obtain a record of that performance. Other tasks require the examinee to produce a product. For many of these products, the measurement of task performance is obtained from an inspection of the tangible end-product itself. The distinction between product and process measurement is clarified in Table 2.

Place Table 2 about here

Developing scoring procedures involves determining the most appropriate way to record the adequacy of the product, processes or both. Following are guidelines in the form of discussions of techniques for scoring products as well as processes and various types of rating scales. Based on these guidelines, TPM scoring procedures can be developed and documented.

Product rating. Since the product, unlike the performance, is a tangible object, product rating is more reliable than process rating. In developing a scoring procedure for a product, all scorable characteristics should be described as accurately and specifically as possible. Those specific characteristics that distinguish a satisfactory product from an unsatisfactory one need to be identified. When the product is to be measured by some type of instrument, the characteristics to be measured must be provided to the examiner along with specific instructions for making the measurement. On the basis of the standard developed it should be possible to decide what is a passing score.

Process rating. An effective scoring procedure for process measurement provides the examiner with a detailed step-by-step description of the process by which the TPM is performed. The task elements/steps and required standards usually are prepared in checklist form. Item 7A on Figure 1 shows a suggested performance checklist format for process rating.

Each task element/step on a performance checklist should be labeled when the sequence is important and when the step is critical. This is done by placing an (S) before steps which must be performed only after the preceding step(s) and a (C) before critical steps where failure to meet standards would result in test failure.

Regardless of whether the output of tested task performance is a product or process, some form of rating scale is generally used. The following sections discuss some of these scales.

Checklists. A checklist is useful for rating the ability to perform a set of procedural steps. Figure 3 shows a portion of a checklist used for rating instrument flying proficiency in an objective and meaningful way.

Place Figure 3 about here

The examiner marks this checklist to record whether each step was performed satisfactorily or unsatisfactorily. Breaking the performance into observable task elements/steps greatly reduces error.

Reliability is usually high in checklist rating because of the nature of the decisions required and the specificity of the procedural steps listed. The more specific the steps on the checklist and the longer the list, the higher the reliability.

Rating scales. A numerical scale divides performance into a fixed number of points. The number of points on the scale depends primarily on the ability of examiners to differentiate between them. Most individuals are able to make at least five differentiations; but even trained examiners can reliably make only about nine. As a result, most numerical scales contain five to nine points.

The descriptive scale uses words or phrases to indicate levels of ability. Figure 4 shows a descriptive scale for rating pilot landing. It uses five different words to describe levels of ability.

Place Figure 4 about here

The degrees of performance can vary to suit the occasion. For example, suppose an operations supervisor wants to evaluate the flying ability of several pilots. The supervisor feels that all the pilots satisfy performance requirements, but there is a need to know the degree to which each is better than satisfactory. By using a descriptive scale, the operations supervisor provides the examiner with a frame of reference. Here the lowest rating possible is labeled "acceptable."

The matter of semantics is a disadvantage in using descriptive scales. An "excellent pilot" does not mean the same thing to all examiners. Another disadvantage is the difficulty in selecting words and phrases which describe degrees of performance that are equally spaced. When the scale shown in Figure 4 is used, some may feel that there is less distance between "excellent" and "superior," than between "fair" and "good."

The graphic scale is a combination of the numerical and descriptive scales. In addition to a numerical scale consisting of a number of points, various words or phrases are set below a continuous horizontal line. The line represents the range of the ability or trait being measured. In using the graphic scale, the examiner must consider both the numerical range of the scale and the words or phrases that describe the various points on the scale.

Three typical forms of a graphic scale are shown in Figure 5. In Example A, instructions for judging the trait of "industry" are provided. The examiner is told to mark the scale after considering "energy and

application to duties, day in and day out." These instructions help reduce errors and improve objectivity and reliability. They also encourage examiners to consider the same traits about each individual.

Place Figure 5 about here

Example B shows a graphic scale in which types of behavior are described for each point on the scale. With most scales, the examiner must not only observe, but also evaluate the performance in order to form a rating. It is generally accepted that people can observe more accurately than they can evaluate what they have seen. Whenever ratings can be based on observations alone, reliability is greatly improved. In Example B the examiner is required only to record, not evaluate, the type of behavior. Hence, this type of graphic scale incorporates much objectivity. In preparing this type of scale, the behavior described for each point must actually be an improvement over the previous point. In addition distances between the points should appear to be about equal.

The scale in Example C is similar to that in B, except descriptive phrases are not provided for all points. Examiners often feel that the rating should fall somewhere between two points. Such a rating is possible with this form of graphic scale. The fuller descriptions in Example C increase the likelihood that observed behavior can be pinpointed on the scale. Generally, the more detailed descriptions should contribute to better rating results.

Establishing cut-off scores. Many tests are designed to assess proficiency on a given task. Since it is often impractical to insist on perfect test scores, it may be necessary to decide upon a cut-off point (a score below which is considered failing or "no-go"). The more complex the skills assessed by the test and the more varied the type of procedure or product, the greater the danger of misclassification (designating a "no-go" as a "go" or vice versa).

Establishing cut-off points is a complex matter. Decisions should be reached only after careful consideration of the general acceptable performance standards for the task(s) and the consequences. A good guideline is that when the consequences of passing one incompetent individual are severe, the cut-off point should be set high. In general, cut-offs are useful when:

1. Mastery of the task is not expected but a suitable level of performance can be specified.
2. Mastery is possible but factors other than competence affect the score (such as careless errors, etc.).

Validation

No matter how carefully the TPM was constructed, it is likely that examinees will encounter problems with the test and testing situation. Thus the TPM must be tried out to identify major deficiencies and to enable examinees to contribute suggestions regarding the elimination of these problems. A validated TPM is one that has been tried out and found to have good predictive validity. That is, the examinees who pass the test are those who can perform the task on the job.

TPMs ought to be validated on the job if resources permit. When this is done, validation is similar to a job analysis performed using the observation technique. A job incumbent is observed performing the task and appropriate entries are made on the TPM checklist or rating scale. The number of incumbents needed for validation is small for unitary tasks and higher for multiple tasks.

There are certain tasks which are rarely performed. Others may be hazardous both to the job incumbent and the observer. In situations where constraints such as these exist, TPMs should be "verified" rather than validated. Subject matter experts are asked to verify the attainment standards, add any missing performance conditions, task elements or steps and delete superfluous ones. They also ensure that the TPM is technically accurate and consistent with current procedures.

Tryout and revise. The following activities ought to be completed in preparation for conducting the tryout.

1. Write administrative instructions for the examiner and examinee. They should be complete, clear and easily understood.
2. Set the TPM aside for a time and then edit it. This practice makes it easier to spot problems which can be corrected before the tryout.
3. Ensure, through the use of subject matter experts and job incumbents, that the items are congruent with the corresponding tasks.
4. Collect all the items of equipment, tools, materials, etc. for administering the TPM.
5. Select job incumbents to serve as examinees.

6. Prepare forms for recording the information to be gathered by observers and/or examiners.
7. Select observers. Observers will watch the administration of the TPM and record information that can be used to correct deficiencies in the items or the assessment procedures themselves. (They may be used as examiners as well as observers.)
8. Provide all necessary information and/or training to the observers and examiners.

As a rule, the tryout should be conducted in steps. First, two to six volunteer job incumbents take the test. The TPM is then revised to correct all the deficiencies identified. Next, additional incumbents are tested; they will be a check on the success (or lack of success) of the revisions made previously. In addition, this second trial may uncover further difficulties. This cycle of testing and revision should be continued until all major deficiencies are corrected.

The cooperating job incumbents ought to be briefed at the outset that the purpose of the tryout is to help identify deficiencies, not to test their proficiency. Furthermore, once difficulties have been identified the incumbents can help develop revisions which will eliminate the problems.

The following are suggested guidelines for obtaining information concerning problems with the TPM.

1. See if the instructions are clear and easily understood by asking the examinees to repeat them in their own words. Note any significant deviations.

2. Record questions asked by the examinees. Written instructions can be prepared to answer frequently asked questions.
3. Record any shortage of supplies or breakdown of equipment.
4. Note ways in which the layout of equipment can be improved without impairing the validity of the TPM.
5. Record all injuries and damage to equipment.
6. Record the length of time taken to complete the TPM.
7. Note any problems experienced in re-establishing a test station for the next examinee.
8. If the test is given in a series of stations, observe any problems experienced in maintaining a smooth flow of examinees from one test station to the next test station.
9. Note any testing conditions which may invalidate the results.
10. When examinees make errors, question them. Ascertain if their errors indicate a misunderstanding of the item.
11. Note any actions of the examiner which might give away the correct answers or confuse the examinee.
12. Score the TPM. Errors provide the basis for a group debriefing. During the debriefing, consensus can be obtained as to what the problem is and what to do about it in terms of a revision.

Record the information collected from the examinees, examiner and observer in sufficient detail to provide a basis for correcting deficiencies. The purpose of the tryout is to make the TPM as valid and reliable as possible. When validating the TPM in the job environment it is also important to verify the conditions, cues, and standards.

A Comment

While worthwhile it should be obvious at this point that considerable effort is involved in constructing and using TPMs. The uninformed instructor may argue against their use because they take considerable time to administer, especially when each trainee must be tested separately. However, this problem can be avoided, while simultaneously enhancing program accountability, by having external examiners do the testing. Craft advisory committee members, retired subject matter experts, potential employers, and other volunteers can effectively augment the program by serving as examiners or observers. This is possible because the TPM was developed in such a way that test administration need no longer be the exclusive responsibility of the instructor.

REFERENCES

AFM 50-2, instructional systems development. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, July 1975.

AFM 50-2, instructional systems development. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, May 1979.

AFP 50-58, volume I introduction, handbook for designers of instructional systems. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, July 1978.

AFP 50-58, volume II task analysis, handbook for designers of instructional systems. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, July 1978.

AFP 50-58, volume III objectives and tests, handbook for designers of instructional systems. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, July 1978.

AFP 50-58, volume IV planning, developing and validating instruction, handbook for designers of instructional systems. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, July 1978.

AFP 50-58, volume V conducting and evaluating instruction, handbook for designers of instructional systems. Headquarters U.S. Air Force, Washington, D.C.: Department of the Air Force, July 1978.

Application of a systems approach to training: a case study (Training Systems and Technology Series: No. 11). Washington, D.C.: U.S. Civil Service Commission, Bureau of Training, June 1969.

Braby, R., Henry, J.M., Parrish, W.F., & Swope. W.M. A technique for choosing cost-effective instructional delivery systems (TAEG Rep. No. 16). Orlando, Florida: Training Analysis and Evaluation Group, Chief of Naval Education and Training, October 1978.

CNTT-A10, procedures for the planning, design, development, and management of navy technical training courses. Naval Air Station, Memphis, Millington, Tennessee: Chief of Naval Technical Training, April 1976.

Curriculum writer's guide (A Handbook for the Development of ARAMCO Vocational Training Literature). Dhahran, Saudi Arabia: ARAMCO Training Department, July 1980.

Handbook for implementation (Interservice procedures for instructional systems development). Washington, D.C.: Naval Health Sciences Education and Training Command, January 1979.

Principles of training systems development (Manual INPO 85-006). Atlanta, Georgia: Institute of Nuclear Power Operations, February 1985.

MIL-STD 1379A, contract training programs. Washington, D.C.: Department of Defense, October 1976.

NAVEDTRA 106A, executive summary and model, interservice procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, August 1975.

NAVEDTRA 106A, phase I analyze, interservice procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, August 1975.

NAVEDTRA 106A, phase II design, interservice procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, August 1975.

NAVEDTRA 106A, phase III develop, interservice procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, August 1975.

NAVEDTRA 106A, phase IV & V implement and control, interservice procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, August 1975.

NAVEDTRA 110, procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, July 1978.

NAVEDTRA 110A, procedures for instructional systems development. Naval Air Station, Pensacola, Florida: Chief of Naval Education and Training, September 1981.

Vineberg, R. & Joyner, J.N. Instructional systems development (ISD) in the armed services: methodology and application (Hum RRO Tech. Rep. 80-1). Alexandria, Virginia: Human Resources Research Organization, January 1980.

APPENDIX

CHEMISTRY TECHNICIAN
TASK PERFORMANCE MEASURE

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Task Title: PERFORM TRITIUM ANALYSIS ON LIQUID SAMPLE Task Number: 6030080111

Trainee: _____ Examiner: _____

Score: ___ PASS ___FAIL Examiner's Signature _____ Date _____

Directions to Trainee:

When I tell you to begin, you are to PERFORM TRITIUM ANALYSIS ON LIQUID SAMPLE. I will describe general conditions and provide the tools to perform this task. Before starting, I will state the task standard(s) as well as initiating cues and answer any questions you have. Performing a critical step improperly or out of sequence will result in test failure.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Conditions:

01 SAMPLE HAS BEEN COLLECTED AND DOCUMENTED

Tools and Equipment:

01 BOILING BEADS
02 CHEMISTRY REPORT
03 DEMINERALIZED WATER
04 DISPOSABLE PAPER TOWELS
05 DISTILLATION APPARATUS
06 FUME HOOD
07 HAND-HELD CALCULATOR
08 IODINE CARRIER SOLUTION
09 LABORATORY COAT
10 LATEX GLOVES
11 LIQUID SCINTILLATION VIALS
12 PACKARD 300 CD SCINTILLATION COUNTER
13 PIPET
14 SAFETY GLASSES
15 SAMPLE TO BE ANALYZED
16 SCINTILLATION SOLUTION
17 SCINTILLATION SPECTROMETER

References:

01 ADMINISTRATIVE PROCEDURES
02 ANALYTICAL PROCEDURE

Attainment Standards:

01 TRITIUM ANALYSIS PERFORMED IN ACCORDANCE WITH ANALYTICAL PROCEDURE

Initiating Cue(s):

01 SAMPLING/SURVEILLANCE SCHEDULE REQUIRED ANALYSIS
02 DIRECTED BY CHEMISTRY SUPERVISOR TO PERFORM ANALYSIS

CHEMISTRY TECHNICIAN
TASK PERFORMANCE MEASURE

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Task Title: PERFORM TRITIUM ANALYSIS ON LIQUID SAMPLE Task Number: 6030080111
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Performance Checklist:	Standards:	Initials:
1 (C) SET UP PACKARD 300 CD SCINTILLATION COUNTER	Steps 9.1-9.17 in AP 1604.004 followed verbatim.	_____
2 (C) DISTILLED SAMPLE (IF REQUIRED)	Distillate colorless upon visual inspection.	_____
3 (C) MEASURED SAMPLE INTO LIQUID SCINTILLATION VIALS	10 ± 0.1 ml pipetted.	_____
4 (C) MEASURED SCINTILLATOR SOLUTION INTO SAMPLE	13 ± 0.1 ml pipetted.	_____
5 (C) MIXED SAMPLE SOLUTION	Vigorously shaken for at least five seconds.	_____
6 (C) MEASURED SAMPLE COUNT RATE	Steps 9.19 and 9.20 in AP 1604.004 followed verbatim. Count rate agreed with known value ± 5%.	_____
7 (C) CALCULATED SAMPLE TRITIUM ACTIVITY	Calculation performed without error. Calculated activity agrees with known value + 5%.	_____
8 RECORDED VALUE OF TRITIUM ACTIVITY	Activity recorded without error.	_____
9 CLEANED AND STORED EQUIPMENT	Glassware washed, rinsed, and returned to proper location.	_____
10 RETURNED REFERENCES	All references returned to proper storage location.	_____

Comments:

Note. This TPM was adapted from Institute of Nuclear Power Operations, Principles of Training System Development, Student Guide, pages DV-19 & 20.

TASK PERFORMANCE MEASURE WORKSHEET

1. JOB TITLE _____ TASK NO. _____

2. TASK STATEMENT (Behavioral action to be taken) _____

_____

3. PERFORMANCE CONDITIONS (Identify those conditions under which the test is to be administered that would make a difference in performance)

   A. SETTING: _____

   B. ____ ACTUAL EQUIPMENT UNDER ACTUAL WORK CONDITIONS

      ____ ACTUAL EQUIPMENT UNDER SIMULATED WORK CONDITIONS

      ____ SIMULATOR, TRAINER, MOCKUP, OR JOB AID TO BE USED

      _____

   C. PERSONNEL REQD: _____

   D. REFERENCES: _____

   E. TOOLS: _____

   F. SUPPORT MATERIALS: _____

   G. TEST EQUIPMENT: _____

   H. SUPPORT EQUIPMENT: _____

   I. OTHER: _____

4. ATTAINMENT STANDARDS (A measure of the adequacy of performance)

   A. 100% OF ALL CRITICAL SCORABLE CHARACTERISTICS/TASK ELEMENTS/STEPS

   B. ___% OF ALL NONCRITICAL SCORABLE CHARACTERISTICS/TASK ELEMENTS/STEPS

   C. TIME REQUIREMENT/LIMIT: _____

   D. OTHER: _____

5. ADMINISTRATIVE INSTRUCTIONS

   A. DIRECTIONS FOR EXAMINER: _____

      _____

   B. DIRECTIONS TO TRAINEE: _____

      _____

6. SCORING PROCEDURES: _____

_____

7. PERFORMANCE MEASURE CHECKLIST (When appropriate (a) place an (S) or (C) to the left of task elements/steps and (b) include a performance rating scale or space for initials)

(S) IMPORTANT SEQUENCE. THIS STEP MUST BE PERFORMED ONLY AFTER THE PRECEDING STEP(S).

(C) CRITICAL STEP. FAILURE TO MEET STANDARDS FOR THIS STEP RESULTS IN TEST FAILURE.

A. PROCESS: _____

TASK ELEMENTS/STEPS	STANDARDS	SCALE/INITIALS

B. PRODUCT: _____

SCORABLE CHARACTERISTICS	STANDARDS	SCALE/INITIALS

Figure 1

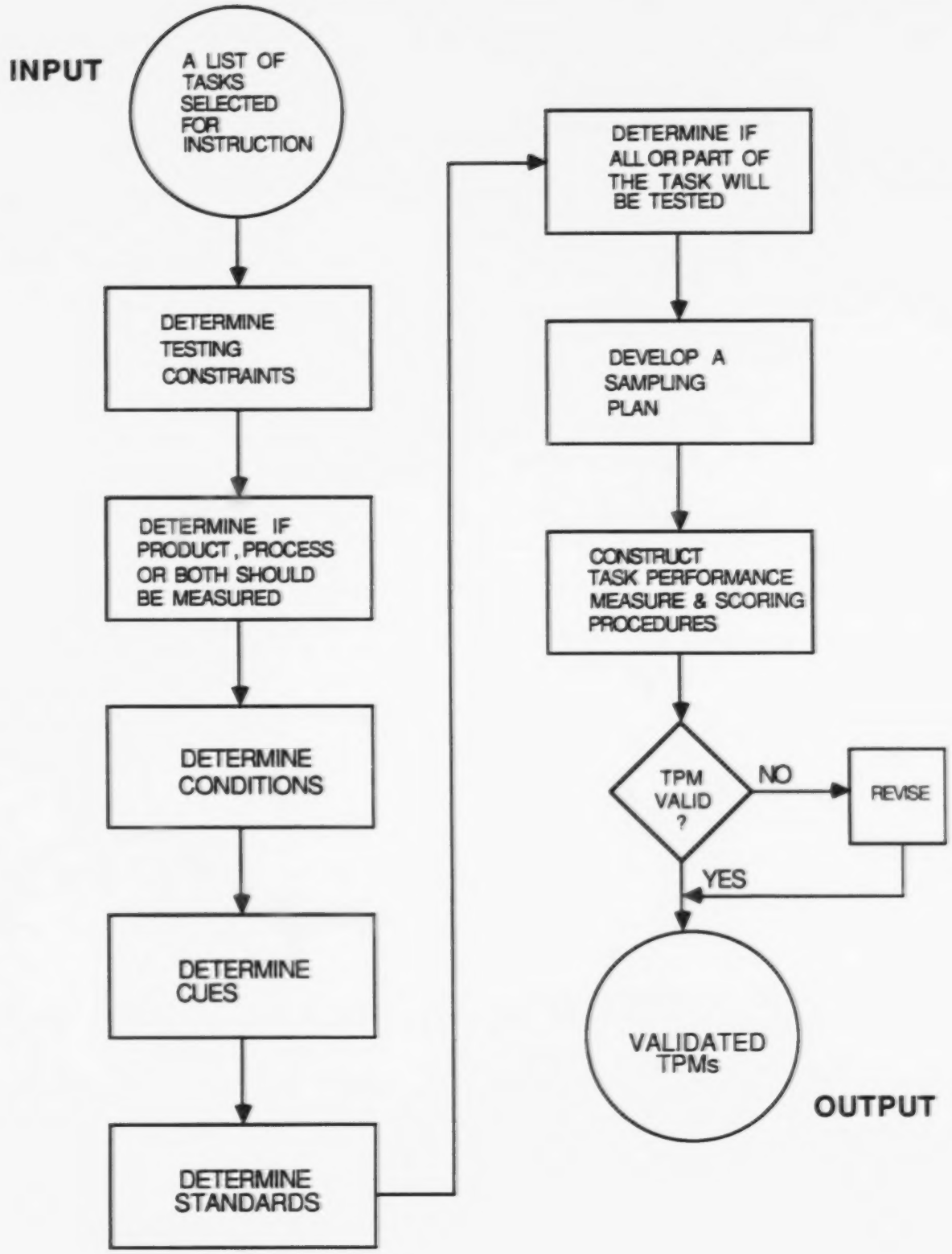

FIGURE 2

Developing Task Performance Measures

Table 1

Task Performance Measure Documentation

JOB TITLE Nurse DUTY/CODE First Aid DATE 1 June 1986 PAGE NO. Page 1 of 2

ITEM CODE	TASK, ELEMENTS T.P.M.	CONDITIONS	INITIATING CUES	STANDARDS
A-001	Perform mouth-to-mouth resuscitation.	CPR manikin used as victim. Victim (manikin) lying face down on the floor.	Examiner says: "You walk into this room and find victim lying face down on the floor. You are alone. Proceed with the actions you deem necessary."	Successful complete each step in sequence.

Procedural Steps

Step 1 OVERVIEW THE SCENE.
Look for indicating signs that might have caused the victim to be lying on floor. None noted. Note by body position and any other evidence if neck fracture might be suspected. None noted.

Step 2 DETERMINE THE CONSCIOUSNESS LEVEL.
Gently shake the victim by the shoulder and in a loud voice (shout) try to arouse: No response from victim, rescuer shouts "help."

Step 3 POSITION VICTIM.
A. Gently straighten victim by placing arm on the side to which the victim is to be turned along the side of the victim's head. Place the victim's other arm along the side of the victim toward the feet. Straighten victim's feet and legs.
B. Supporting the victim's head and neck with the head hand, use the foot hand to roll the victim as a unit toward the rescuer, until victim is on back. Move the arm of victim that was supporting head to along the victim's side.

Step 4 TILT HEAD.
Place foot hand under victim's neck and head hand on victim's forehead and extend airway by tilting forehead back and jutting chin upwards.

TABLE 1
(Continued)

JOB TITLE Nurse DUTY/CODE First Aid DATE 1 June 1986 PAGE NO. Page 2 of 2

ITEM CODE	TASK, ELEMENTS T.P.M.	CONDITIONS	INITIATING CUES	STANDARDS

Step 5 LOOK, LISTEN AND FEEL.
Rescuer leans over victim placing ear over victim's mouth & nose, facing feet of victim.

A. Look at victim's chest to check for rising and falling indication of breathing.
B. Listen at victim's mouth and nose to hear breathing.
C. Feel at victim's mouth and nose for breath exhalations on rescuer's cheek and ear.

Step 6 PINCH NOSE SHUT.
Rescuer, with head hand, pinches victim's nostrils shut to prevent air leakage. Foot hand is under victim's neck to hold position of neck extension to maintain open airway.

Step 7 SEAL MOUTH, GIVE FOUR QUICK BREATHS.

A. Rescuer seals victim's mouth with his mouth and gives four quick full breaths.
B. Rescuer feels breaths going into victim's lungs and observes victim's chest rising & falling with breaths.

Step 8 CHECK CAROTID PULSE.
Rescuer pauses and checks victim's carotid pulse for at least five seconds. Using index and middle fingers, rescuer finds victim's trachea and slides fingers toward the rescuer into the groove of the neck where carotid artery may be palpated.

Examiner says: "You have a pulse but no breathing."

Step 9 CONTINUE RESCUE BREATHING.

A. Rescuer pinches victim's nostrils shut with head hand and supports neck with foot hand to maintain open airway.
B. Rescuer seals victim's mouth with his mouth and gives one breath every five seconds (12 per minute).

Examiner says: "Victim is revived, you are now relieved."

END OF TEST

TABLE 2

Performance Measurement

by Means of Product or Process

OUTPUT	EVALUATION	EXAMPLES
	Product	
A product which is tangible and observable.	The product itself displays TPM success or failure. The product can be examined directly and compared to the standards.	(a) Filled out form (b) Machined workpiece (c) Assembled item (d) Typed paragraph (e) Link trainer product of performance
	Process	
A procedure which is observable but transient.	The performance must be observed and (a) scored at the time it occurs, or (b) recorded by some means e.g. video tape for subsequent scoring.	(a) Instructing (b) Pilot landing (c) Safe driving (d) Hiking (e) Apprehending a suspect (f) Mouth-to-mouth resuscitation

Instructions: If the performance was satisfactory place a check mark in the "Sat" column. If the performance was unsatisfactory, place a check mark in the "Unsat" column.

	Sat.	Unsat.
1.(c) Maintained constant heading (within 5^0 of course)	____	____
2.(c) Maintained constant altitude (within 50 feet)	____	____
3.(c) Made a timed turn (gyros caged) (within 10^0 of new heading)	____	____
4.(c) Made a steep turn (within (50 feet of altitude)	____	____

Figure 3 Checklist for Rating Instrument Flying Proficiency

INSTRUCTIONS: Place a check mark in the box above the word that most accurately describes the pilot being rated.

ACCEPTABLE	FAIR	GOOD	EXCELLENT	SUPERIOR

Figure 4 Descriptive Scale for Rating Pilot Landing

Example A

Industry: Consider energy and application to duties day in and day out.

1	2	3	4	5
Lazy	Indifferent	Diligent	Energetic	Untiring

Example B

Cooperation: Demonstration of willingness to work with others.

1	2	3	4	5
Creates friction	Indifferent to others	Gets along with most people	A harmonious team worker	Actively promotes harmony in working with others

Example C

Initiative: Action taken on own responsibility.

1	2	3	4	5
Slow to act, even when a decision is needed. Waits for others. Lets opportunities pass. Does not volunteer. Reticent.		Takes needed action without delay. Volunteers for some tasks. Undertakes all routine jobs without supervision. Dependable.		Anticipates needs. Works ahead and prepares for possibilities. Actively seeks opportunities. Eager.

FIGURE 5 Forms of a Graphic Scale